A Fully Convolutional Neural Network for Comprehensive Compartmentalization of Abdominal Adipose Tissue Compartments in MRI

Yeshe M. Kway1, 2, Kashthuri Thirumurugan1, Navin Michael1, Kok Hian Tan3,4, Keith M Godfrey5, Peter Gluckman1, Yap Seng Chong1, 6, Kavita Venkataraman7, Eric Yin Hao Khoo2, Chin Meng Khoo2, 8, Melvin Khee-Shing Leow1, 9, 10, 11, E Shyong Tai2, 12, Jerry KY. Chan13, 14, Shiao-Yng Chan1, 6, Johan G. Eriksson1, 6, 15, 16, Marielle V. Fortier1, 17, Yung Seng Lee1, 18, 19, S. Sendhil Velan1, Mengling Feng7, 20,\*, Suresh Anand Sadananthan1,\*

1Singapore Institute for Clinical Sciences, Agency for Science Technology, and Research, Singapore.

2Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore.

3Duke-National University of Singapore Graduate Medical School, Singapore.

4Department of Maternal Fetal Medicine, KK Women’s and Children’s Hospital, Singapore

5MRC Lifecourse Epidemiology Centre & NIHR Southampton Biomedical Research Centre, University of Southampton & University Hospital Southampton NHS Foundation Trust, Southampton, United Kingdom.

6Department of Obstetrics and Gynaecology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore.

7Saw Swee Hock School of Public Health, National University of Singapore, National University Health System, Singapore.

8Department of Medicine, National University Health System, Singapore.

9Lee Kong Chian School of Medicine, Nanyang Technological University (NTU), Singapore.

10Human Potential Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore.

11Department of Endocrinology, Division of Medicine, Tan Tock Seng Hospital (TTSH), Singapore.

12Division of Endocrinology, University Medicine Cluster, National University Health System, Singapore

13Department of Reproductive Medicine, KK Women's and Children's Hospital, Singapore.

14Experimental Fetal Medicine Group, Department of Obstetrics and Gynaecology, Yong Loo Lin School of Medicine, National University Health System, Singapore.

15Department of General Practice and Primary Health Care, University of Helsinki and Helsinki University Hospital, Helsinki, Finland.

16Folkhälsan Research Center, Helsinki, Finland.

17Department of Diagnostic and Interventional Imaging, KK Women’s and Children’s Hospital, Singapore.

18Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Republic of Singapore

19Division of Paediatric Endocrinology, Department of Paediatrics, Khoo Teck Puat-National University Children’s Medical Institute, National University Hospital, National University Health System, Singapore, Republic of Singapore.

20Institute of Data Science, National University of Singapore, Singapore.

\* M.F. and S.A.S. are co-senior authors.

**Corresponding Author:**

Suresh Anand Sadananthan

Email: [suresh@sics.a-star.edu.sg](mailto:kwayy@sics.a-star.edu.sg)

Tel.: +65 64070567

Address: Singapore Institute for Clinical Sciences, Agency for Science Technology and Research, 30 Medical Drive, Singapore 117609**.**

**Abstract**

**Background:** Existingliterature has highlighted structural, physiological, and pathological disparities among abdominal adipose tissue (AAT) sub-depots. Accurate separation and quantification of these sub-depots are crucial for advancing our understanding of obesity and its comorbidities. However, the absence of clear boundaries between the sub-depots in medical imaging data has challenged their separation, particularly for internal adipose tissue (IAT) sub-depots. To date,thequantification of AAT sub-depots remains challenging, marked by a time-consuming, costly, and complex process.

**Purpose:** To implement and evaluate a convolutional neural network to enable granular assessment of AAT by compartmentalization of subcutaneous adipose tissue (SAT) into superficial subcutaneous (SSAT) and deep subcutaneous (DSAT) adipose tissue, and IAT into intraperitoneal (IPAT), retroperitoneal (RPAT), and paraspinal (PSAT) adipose tissue.

**Material and Methods:** MRI datasets were retrospectively collected from Singapore Preconception Study for Long-Term Maternal and Child Outcomes (S-PRESTO: 389 women aged 31.4 ± 3.9 years) and Singapore Adult Metabolism Study (SAMS: 50 men aged 28.7 ± 5.7 years). For all datasets, ground truth segmentation masks were created through manual segmentation. A Res-Net based 3D-UNet was trained and evaluated via 5-fold cross-validation on S-PRESTO data (N=300). The model’s final performance was assessed on a hold-out (N=89) and an external test set (N=50, SAMS).

**Results:** The proposed method enabled reliable segmentation of individual AAT sub-depots in 3D MRI volumes with high mean Dice similarity scores of 98.3%, 97.2%, 96.5%, 96.3%, and 95.9% for SSAT, DSAT, IPAT, RPAT, and PSAT respectively.

**Conclusion:** Convolutional neural networks can accurately sub-divide abdominalSAT into SSAT and DSAT, and abdominal IAT into IPAT, RPAT, and PSAT with high accuracy. The presented method has the potential to significantly contribute to advancements in the field of obesity imaging and precision medicine.

**Key words:** Water-fat MRI, Abdominal fat segmentation, Deep learning, Convolutional neural network

# Introduction

Abdominal obesity is a key independent factor for the development of obesity-related disorders, including type 2 diabetes (T2D) and cardiovascular disease (CVD), and is characterized by excess accumulation of abdominal adipose tissue (AAT) (Shuster et al., 2012). AAT is a highly heterogeneous tissue; distinct AAT depots show different associations with cardiometabolic risk factors (Lee et al., 2013; Kwok et al., 2016). AAT is generally classified into subcutaneous adipose tissue (SAT) and internal adipose tissue (IAT). SAT can be further separated into deep (DSAT) and superficial (SSAT) subcutaneous adipose tissue, by the facia superficialis, and IAT can be further separated based on anatomical spaces (Shen et al., 2003). IAT within the intraperitoneal (intraperitoneal adipose tissue (IPAT)) and retroperitoneal (retroperitoneal adipose tissue (RPAT)) space constitute the visceral adipose tissue (VAT), while IAT within the paraspinal space is defined as paraspinal adipose tissue (PSAT). Because separation of distinct AAT sub-depots in medical images is challenging, studies have predominantly quantified SAT as a single adipose tissue depot and VAT as IAT (Shen et al., 2003; Hu et al., 2016; Borga, 2018). However, all distinct AAT sub-depots exhibit differences in their cellular structure, morphology, metabolic activity, and association with cardiometabolic diseases (Lee et al., 2013; Kwok et al., 2016). Therefore, an automated and standardized method for accurate quantification of individual AAT sub-depots can provide a better understanding of obesity and its physiological and pathological phenotypes.

VAT shows increased lipolytic activity and pro-inflammatory characteristics compared to SAT and is strongly linked to the development of T2D and CVD (Freedland, 2004; Ibrahim, 2010; Shuster et al., 2012; Lee et al., 2013; Tchernof and Després, 2013). Though VAT as a whole is a strong indicator of metabolic disease risk, VAT sub-depots have been identified to differ in terms of their metabolic function and association with obesity related risk factors and diseases (Shen et al., 2003; Yang et al., 2008; Item and Konrad, 2012; Anoop et al., 2017; Tanaka et al., 2020, 2021). For example, increased accumulation of RPAT may not be as deleterious compared to accumulation of IPAT (Tanaka et al., 2020, 2021). Interestingly, studies have shown that IPAT exhibits increased lipolysis and proinflammatory cytokine secretion, compared to other AAT depots (Yang et al., 2008). The fact that IPAT (omental and mesenteric adipose tissue) is drained by the portal vein, while RPAT is drained by the inferior vena cava (systemically), motivated the formulation of the portal theory. The portal theory postulates that secretory products derived from IPAT contribute to increased hepatic exposure to free fatty acids and inflammatory cytokines, which subsequently promote the development of steatosis, hepatic insulin resistance, and T2D (Arner, 1998; Item and Konrad, 2012). However,current studies have predominantly considered and quantified VAT as a single entity rather than segmenting it into IPAT and RPAT (Shuster et al., 2012; Kwok et al., 2016). Similar to VAT, SAT sub-depots, DSAT and SSAT have shown distinct associations with cardiometabolic health. DSAT might share similar deleterious characteristics to VAT, while SSAT could be a protective fat storage site (Kelley et al., 2000; Lundbom et al., 2013; Marinou et al., 2014). However, conflicting findings have also been reported by other studies (Miyazaki et al., 2002; Walker et al., 2014).

While VAT and SAT have been extensively studied in the context of white adipose tissue and cardiometabolic disease, RPAT and PSAT have been identified as potential sites of brown adipose tissue (BAT) (Leitner et al., 2017). BAT contributes significantly to thermogenesis and energy expenditure, making it a potential target for obesity treatment (Villarroya et al., 2017). Since data on BAT function and location are limited, accurate segmentation of potential BAT storage sites could improve identification and characterization of BAT. In addition, adipose tissue infiltration into the lumbar paraspinal musculature is observed in neuromuscular disease, chronic lower back pain, and symptomatic lumbar spinal stenosis (Hadar et al., 1983; Yanik et al., 2013; Chen et al., 2014; Kalichman et al., 2017). While little is known about the relationship between PSAT accumulation and fat infiltration into the paraspinal musculature, automated PSAT quantification could help to understand and characterize pathological conditions in neuromuscular and spinal disorders (Zhang et al., 2021). In summary, individual AAT sub-depots exhibit morphological and functional differences, therefore, accurate differentiation and quantification of individual AAT depots can contribute to improved understanding of multiple domains.

Accurate quantification of AAT depots can be achieved by computerized tomography (CT) and magnetic resonance imaging (MRI). Several studies have manually segmented AAT depots from imaging data (Lee et al., 2013; Kwok et al., 2016; Kullberg et al., 2017).However, manual segmentation is expensive, time consuming, and can introduce variability between readers. While conventional image processing methods have been implemented to automatically quantify SAT and VAT (Hu et al., 2016; Kullberg et al., 2017; Borga, 2018)), their translation to segment distinct AAT sub-depots has been hindered by the complex structure and the low-intensity contrast between distinct AAT sub-depots. Unlike conventional image processing methods, fully convolutional neural networks (FCNN) have the potential to address the aforementioned challenges (Long et al., 2014). Several studies have presented FCNN to accurately quantify skeletal muscle, total AAT, SAT, and VAT volumes in MRI and CT (Wang et al., 2017; Weston et al., 2019; Dabiri et al., 2020; Estrada et al., 2020; Küstner et al., 2020). Nonetheless, deep learning applications for AAT segmentation have mainly focused on SAT and VAT quantification; so far, only two studies have explored the use of FCNN to delineate SAT into SSAT and DSAT (Bhanu et al., 2021; Kway et al., 2021). Although several reviews have highlighted the need for more sophisticated methods to quantify AAT sub-depots (Shen et al., 2003; Gantz et al., 2011; Katzmarzyk et al., 2012), to the best of our knowledge, to date, there are no automated algorithms that enable comprehensive volumetric compartmentalization of anatomically distinct AAT sub-depots. In summary, the main contributions of this work are:

* Our study explores the feasibility of employing a fully convolutional neural network to accurately compartmentalize and quantify abdominal adipose tissue sub-depots – specifically, superficial subcutaneous, deep subcutaneous, intraperitoneal, and retroperitoneal adipose tissue depots – in 3-dimensional magnetic resonance imaging data.
* High-quality volumetric ground truth data were created meticulously by experts for a total of 439 imaging volumes, across two different scanners and populations, for effective model training and evaluation.
* Thorough evaluation, comprising 5-fold cross-validation and testing on an external data set was undertaken to assess the performance of our presented model. By achieving high accuracy in quantifying distinct abdominal adipose tissue sub-depots, the model allows for a comprehensive assessment of abdominal obesity and, therefore, presents a significant contribution to the field of obesity imaging.

The rest of the paper is organized as follows. Section 2 provides information on the cohorts, MRI protocol used for data collection, segmentation model, and the metrics used for performance validation. The results of the segmentation model are shown in Section 3 and discussed in detail in Section 4. Lastly, we conclude our work in Section 5.

# Material and Methods

## *Study Population*

Imaging data from participants of Singapore Preconception Study of Long-Term Maternal and Child Outcomes (S-PRESTO) (Loo et al., 2021) were used as training database (N=389). Generalization ability of the proposed model was evaluated on an external test set, the Singapore Adult Metabolism Study (SAMS) (N=50) (Khoo et al., 2014). Both cohort studies included individuals from three different Singaporean ethnic groups, Chinese, Malay, and Indian. All participants gave written consent and studies were approved by the institutional review board. Cohort demographics and imaging details are shown in Table 1.

### *Singapore Preconception Study of Long-Term Maternal and Child Outcomes (S-PRESTO)*

S-PRESTO is a longitudinal birth cohort study which recruited women in reproductive age (18-45 years) between February 2015 and October 2017. Detailed study information can be found in (Loo et al., 2021). 389 participants who underwent abdominal MRI at recruitment were included in the current study. The participants represented a wide range of BMI categories, ranging from underweight to obese.

### *Singapore Adult Metabolism Study (SAMS)*

The SAMS study aimedto investigate the relationship between body fat partitioning and metabolic health, such as insulin resistance in a multi-ethnic cohort of adult Singaporean men aged 21 to 40 years (Khoo et al., 2014). Imaging volumes from 50 participants were randomly selected to create an external test set (Table 1).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 1. Cohort characteristics** | | | | |
| **Characteristics** |  | **S-PRESTO (Women, N=389)** |  | **SAMS (Men, N=50)** |
| **Age, years** |  | 31.4 ± 3.9 |  | 28.7 ± 5.7 |
| **Weight, kg** |  | 59.8 ± 13.1 |  | 71.3 ± 9.0 |
| **Height, cm** |  | 159.6 ± 5.4 |  | 171.8 ± 5.5 |
| **BMI Categories, (kg/m2)**  Underweight (<18.5)  Normal weight (≥18.5, <23)  Overweight (≥23, <27.5)  Obese (≥27.5) |  | 37 (9.6)  183 (47.5)  96 (24.9)  69 (17.9) |  | 3 (6.0)  15 (30.0)  23 (46.0)  9 (18.0) |
| **Ethnicity**  Chinese  Malay  Indian  Mixed |  | 266 (68.7)  70 (18.1)  42 (10.9)  9 (2.3) |  | 26 (52.0)  12 (24.0)  12 (24.0)  N.A. |
| **Imaging information**  Scanner  Sequence  Repetition time (msec)  Echo time 1 (msec)  Echo time 2 (msec)  Flip angle (°)  Image dimension (XY)  Image resolution (mm) |  | Siemens Magnetom Skyra  two-point-Dixon  3.86  1.23  2.46  9  320320  (1.5, 1.5, 3) |  | Siemens Trim Trio  two-point Dixon  5.28  2.45  3.68  9  320320  (1.25,1.25, 3) |
| Data are presented in frequencies (%) or mean ± SD  BMI categories were defined using Asian specific cut-offs | | | | |

## *Magnetic Resonance Imaging*

Axial MR images of the abdomen were acquired using 2-point Dixon sequence and body matrix coil during a breath-hold of 16-18 s on Siemens 3T Skyra scanner (S-PRESTO) and on Siemens 3T Tim Trio scanner (SAMS). To standardize the volumetric assessment, all imaging volumes were restricted to range from L1 to L5 lumbar vertebrae. In S-PRESTO, two separate imaging volumes were acquired (top of the liver to L3 lumbar vertebrae and L3 lumbar vertebrae to upper sacrum) and therefore combined in advance.

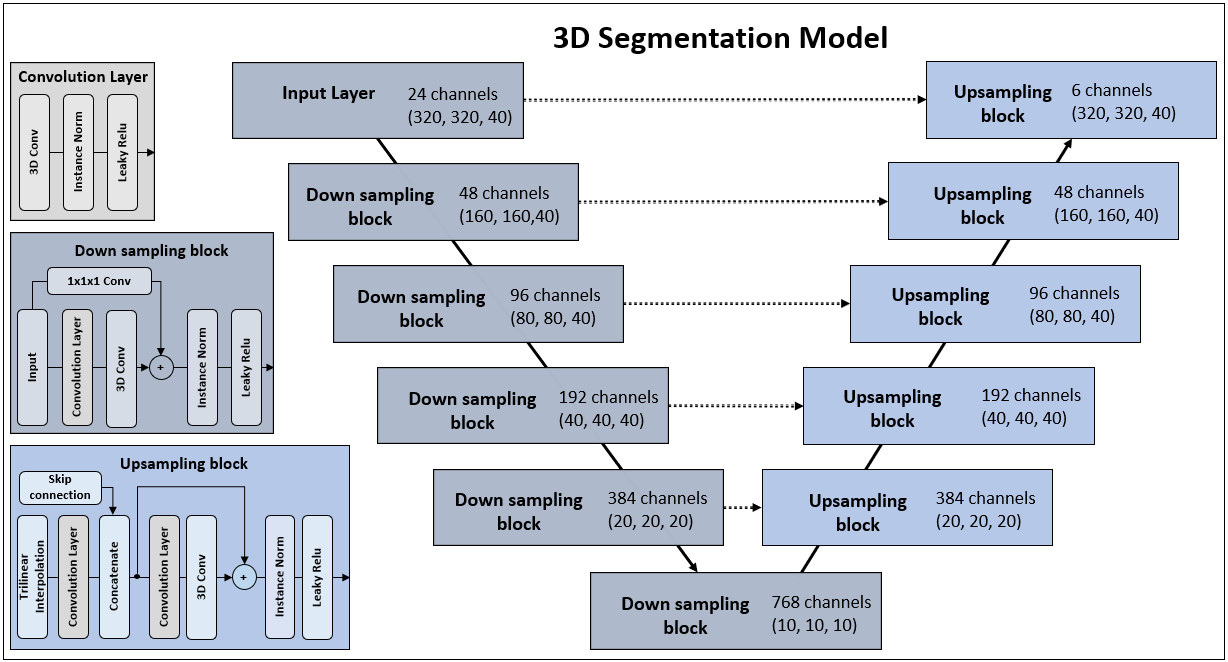
## *Ground Truth Generation*

A graph-theoretical segmentation technique was used to create the initial segmentation masks for total SAT and total intra-AAT (Sadananthan et al., 2019). The resulting masks were then manually corrected for mis-segmentations of non-adipose tissue voxels. Subsequently, SAT was delineated into SSAT and DSAT by tracing the facia superficialis and VAT voxels were manually classified into IPAT, RPAT, or PSAT, based on their anatomical location. Taxonomy was adopted from (Shen et al., 2003) to follow standardized nomenclature and adipose tissue topography. Fat voxels within the intraperitoneal space and retroperitoneal space were labelled as IPAT and RPAT, respectively. Fat voxels within the paraspinal space were labelled as PSAT (paravertebral adipose tissue and intermuscular adipose tissue). The landmarks to assess IPAT and RPAT were previously assessed by dissection in human cadavers and reported as highly accurate for volumetric quantification (Abate et al., 1994). Anatomical boundaries, including the ascending and descending colon, abdominal aorta, inferior vena cava, kidneys, psoas muscles, vertebral bodies, and liver were used to separate the anatomical spaces from each other. A visual example of delineation of the relevant anatomical spaces and a detailed list of the landmarks used can be found in Figure S1 [supplement]. Manual refinement and segmentation were performed by a trained MR reader (YMK) with 4 years of experience in MRI, under supervision of a radiologist (MVF) with > 20 years of experience, using ITKSnap software ([www.itksnap.org](http://www.itksnap.org)). The volumes were calculated by multiplying the sum of labelled voxels of a respective fat depot, with the voxel resolution. Intra-reader reliability assessed for repeated segmentation of 20 randomly selected volumes resulted in Dice similarity coefficients of 0.98, 0.96, 0.94, 0.93, and 0.94 for SSAT, DSAT, IPAT, RPAT, and PSAT respectively.

## *Segmentation Model*

A 3D-UNet (Ronneberger et al., 2015), comprised of ResNet blocks (He et al., 2015), was designed based on implementation methods presented in (Isensee et al., 2018). Down sampling was performed with stride operations within the encoder, trilinear interpolation was used to up-sample feature maps within the decoder, and the leaky rectified linear unit (Leaky-ReLU) transfer function was used as activation function for all layers (Xu et al., 2015). A detailed model description is illustrated in Figure 1. Glorot uniform initialization was used for weight initialization (Xavier Glorot and Yoshua Bengio, 2010). Hyper-parameters, including batch size, learning rate, and patience were determined on a preliminary experiment, empirically. The final model configuration was trained and evaluated in a 5-fold cross-validation experiment, defined with a batch size of two, learning rate of, and patience of 40. The Adam optimizer (Kingma and Ba, 2014) was used to minimize the loss function, which was defined as a label-wise summation of the binary cross-entropy and the generalized dice loss:

**Figure 1. Model Architecture:** The model is comprised of 11 building blocks, containing 59,145,102 trainable parameters. Numbers within the blocks indicate dimension (x, y, z). Operations within the input layer are the same as those performed in the down sampling block with the only difference that no downsampling operations are performed. Downsampling is performed with a stride operation of 2 in the first convolutional layer in each down sampling block. The network is designed to first pool over the x- and y-axis until their dimensions match with the z-axis; after that, all axes are downsampled synchronously. The number of convolutional filters is doubled after each downsampling block, starting with 24 filters in the first block.



|  |  |
| --- | --- |
|  | (1) |

Here, is the probability matrix for class c and is the corresponding ground truth matrix. The model was implemented using PyTorch (v. 1.9.0) and trained on an Nvidia Tesla V100 graphics processing unit.

## *Preprocessing and Data Augmentation*

Imaging volumes were normalized to a scale of 0–1. Patches with a shape of were cropped from each image volume randomly within each training cycle. Data augmentation was implemented using the TorchIO library (Pérez-García et al., 2021). Figure S2 [supplement] provides a visualization of the customized augmentation pipeline implemented in this study. The augmentation pipeline incorporated a stochastic element where, for each time a volume was loaded into memory, there was a 50% chance of augmentation being applied. Augmentation included the addition of noise, affine and elastic transformation, as well as the synthesis of bias field inhomogeneity (van Leemput et al., 1999), ghosting (to simulate respiratory and cardiac motion), and motion artifacts (to mimic patient movements) (Richard Shaw et al., 2019).

## *Modelling*

Imaging volumes from the S-PRESTO cohort (N=389) were used for deep learning modelling. To ensure statistical stability of the presented results, the segmentation model was trained and evaluated using 5-fold cross-validation.In advance, a total of 89 volumes were randomly excluded to create a hold-out test set to prevent the chance of data leakages.To assess the robustness and generalizability of the developed model, the model was additionally evaluated on an external test set (SAMS, N=50). All five cross-validation models were trained on 240 MRI volumes and evaluated on their respective cross-validation test set (N=60), the hold-out test set (N=89), and the external test set (N=50). The following work presents the mean performance of all models on all test sets.

## *Statistical Analysis*

The accuracy of the segmentation model was evaluated against manually generated ground truth data. The segmentation approach was quantitatively evaluated using Dice similarity coefficients (Dice) for overlap assessment and the 95th percentile of the Hausdorff Distance (HD95) for contour agreement. Furthermore, false positive (FP) and false negative (FN) rates, precision, and sensitivity were computed to provide a comprehensive evaluation of the segmentation performance. Detailed metric definitions are provided in Table S4 [supplement]. Dependent-sample t-tests were used to assess AAT depot-specific metric differences within the same testing set, while independent-sample t-tests were used to assess AAT depot-specific metric differences across different test sets. Statistical significance was set at a significance level of P < 0.05. To evaluate the volumetric agreement between the predicted and ground truth data, Bland-Altman plots were used to evaluate a randomly selected model from the 5-fold cross-validation experiment, denoted by “model 1” in Section 3.

# Results

## *Mean Performance: 5-Fold Cross-Validation*

Table 2 presents the model’s mean performance across the 5-fold cross-validation experiment for both the hold-out test set (N=89) and the external test set (N=50). Table S1 (supplement) shows the mean performance over the 5-fold cross-validation test sets.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 2. Mean evaluation results on hold-out and external test set** | | | | | | | | | | | | | |
| **Fat Depot** | **Hold-Out Test Set (S-PRESTO, Women, N=89)** | | | | | |  | **External Test Set (SAMS, Men, N=50)** | | | | | |
| **Dice**  **(%)** | **HD95**  **(mm)** | **FP**  **(%)** | **FN**  **(%)** | **Precision**  **(%)** | **Sensitivity**  **(%)** |  | **Dice**  **(%)** | **HD95**  **(mm)** | **FP**  **(%)** | **FN**  **(%)** | **Precision**  **(%)** | **Sensitivity**  **(%)** |
| **SSAT** | 98.21 ± 0.05 | 1.78 ± 0.86 | 1.71 ± 0.20 | 0.19 ± 0.02 | 98.15 ± 0.15 | 98.19 ± 0.16 |  | 95.66 ± 3.61 | 2.07 ± 1.62 | 0.30 ± 0.26 | 4.22 ± 4.18 | 95.66 ± 4.22 | 95.78 ± 4.14 |
| **DSAT** | 97.29 ± 0.07 | 1.80 ± 0.90 | 2.77 ± 0.31 | 0.17 ± 0.02 | 97.38 ± 0.37 | 97.05 ± 0.32 |  | 94.54 ± 3.64 | 3.48 ± 10.41 | 0.21 ± 0.07 | 5.28 ± 3.84 | 94.46 ± 4.42 | 94.72 ± 3.80 |
| **IPAT** | 96.76 ± 0.15 | 1.54 ± 0.24 | 3.37 ± 0.50 | 0.07 ± 0.0002 | 96.91 ± 0.23 | 97.22 ± 0.32 |  | 93.45 ± 4.16 | 2.29 ± 2.62 | 0.15 ± 0.10 | 7.43 ± 4.60 | 94.48 ± 4.89 | 92.57 ± 4.55 |
| **RPAT** | 96.50 ± 0.09 | 1.54 ± 0.25 | 3.52 ± 0.24 | 0.06 ± 0.00 | 96.53 ± 0.26 | 96.73 ± 0.20 |  | 92.71 ± 4.26 | 3.71 ± 4.38 | 0.13 ± 0.07 | 8.64 ± 6.03 | 94.24 ± 2.76 | 91.36 ± 5.97 |
| **PSAT** | 96.08 ± 0.19 | 1.22 ± 0.59 | 4.24 ±  0.40 | 0.02 ± 0.004 | 96.41 ± 0.38 | 95.78 ± 0.39 |  | 88.49 ± 6.02 | 3.52 ± 2.59 | 0.03 ± 0.02 | 14.14 ± 9.14 | 91.88 ± 3.71 | 85.86 ± 9.05 |
| Quantitative evaluation metrics are in %, presented as mean ± standard deviation. S-PRESTO = Singapore’s Preconception Study of Long-Term Maternal and Child Outcomes, SAMS = Singapore Adult Metabolism Study, SSAT = Superficial Subcutaneous Adipose Tissue, DSAT = Deep Subcutaneous Adipose Tissue, IPAT = Intraperitoneal Adipose Tissue, RPAT = Retroperitoneal Adipose Tissue, PSAT = Paraspinal Adipose Tissue, Dice = Dice Similarity Coefficient, FP = False Positive Rate, FN = False Negative Rate. | | | | | | | | | | | | | |

### *Performance on the Hold-out Test Set*

Themean segmentation performance on the hold-out test set resulted in Dice values of over 96% for all individual AAT depots (Dice: SSAT=98.2%, DSAT=97.3%, IPAT=96.8%, RPAT=96.5%, and PSAT=96.1%)) and mimicked the mean performance seen on the cross-validation test sets. In addition, HD95 scores indicated strong contour agreement with the ground truth data, across all AAT depots (Table 2). Among the SAT depots, segmentation accuracy was higher for SSAT (precision and sensitivity of 98.2% and 98.2%, respectively), than for DSAT (precision and sensitivity of 97.4% and 97.1%, respectively; all P<0.05). Compared to the SAT depots, the segmentation accuracy for IAT depots was slightly lower (precision and sensitivity 96.9% and 97.2% for IPAT, 96.5% and 96.7% for RPAT, and 96.4% and 95.8% for PSAT, respectively; all P<0.05). While false positive rates were slightly elevated for IAT depot predictions (false positive rates, (IPAT=3.37%, RPAT=3.52%, PSAT=4.24%) vs (SSAT=1.71%, DSAT=2.77%); all P<0.05), false negative rates were slightly higher for SAT depot predictions (false negative rates, (SSAT=0.19%, DSAT=0.17%) vs (IPAT=0.07%, RPAT=0.06%, PSAT=0.02%), all P<0.05). While estimates for PSAT showed the highest rate of false positive predictions (4.24%), they showed the lowest false negative rates (0.02%).

### *Performance on the External Test Set*

The evaluation on the external test set resulted in slightly reduced segmentation performance, with a reduction in mean Dice of 2.6%, 2.8%, 3.3%, 3.8%, and 7.6% for SSAT, DSAT, IPAT, RPAT, and PSAT, respectively, when compared to the segmentation performance on the hold-out test set (Table 2). This slight performance reduction was also evident in the HD95 metric, with a notable increase in the standard deviation. Overall, the largest performance drop was observed for PSAT (Dice, (hold-out: 96.1%) vs. (external=88.5%), P=0.13), attributed to an increase in false negative prediction (false negatives: (hold-out: 0.02%) vs. (external: 14.14%), P=0.09). The performance for the other AAT depots remained highly robust with mean Dice of 95.7%, 94.5%, 93.5%, and 92.7% for SSAT, DSAT, IPAT, and RPAT, respectively. The implemented augmentation scheme (Figure S2 [supplement]) improved the generalizability of the trained model, to the external test set, by an average increase in Dice of 1.17%, 0.83%, 0.62%, 3.28% for DSAT, IPAT, RPAT, and PSAT respectively; no significant change in performance was found for SSAT (Table S3 [supplement]).

## *Volumetric Quantification Performance*

Median values for predicted and ground truth volumes as well as their percentage differences, with respect to the ground truth volume, are shown in Table S2 [supplement]. Bland-Altman plots, for all AAT depots are shown for the predictions on the S-PRESTO hold-out test set and the external test set (SAMS data) in Figure 2 and 3, respectively. Overall, the plots indicated strong agreement between predicted and ground truth volumes, with almost all estimates falling within the 95% limits of agreement.

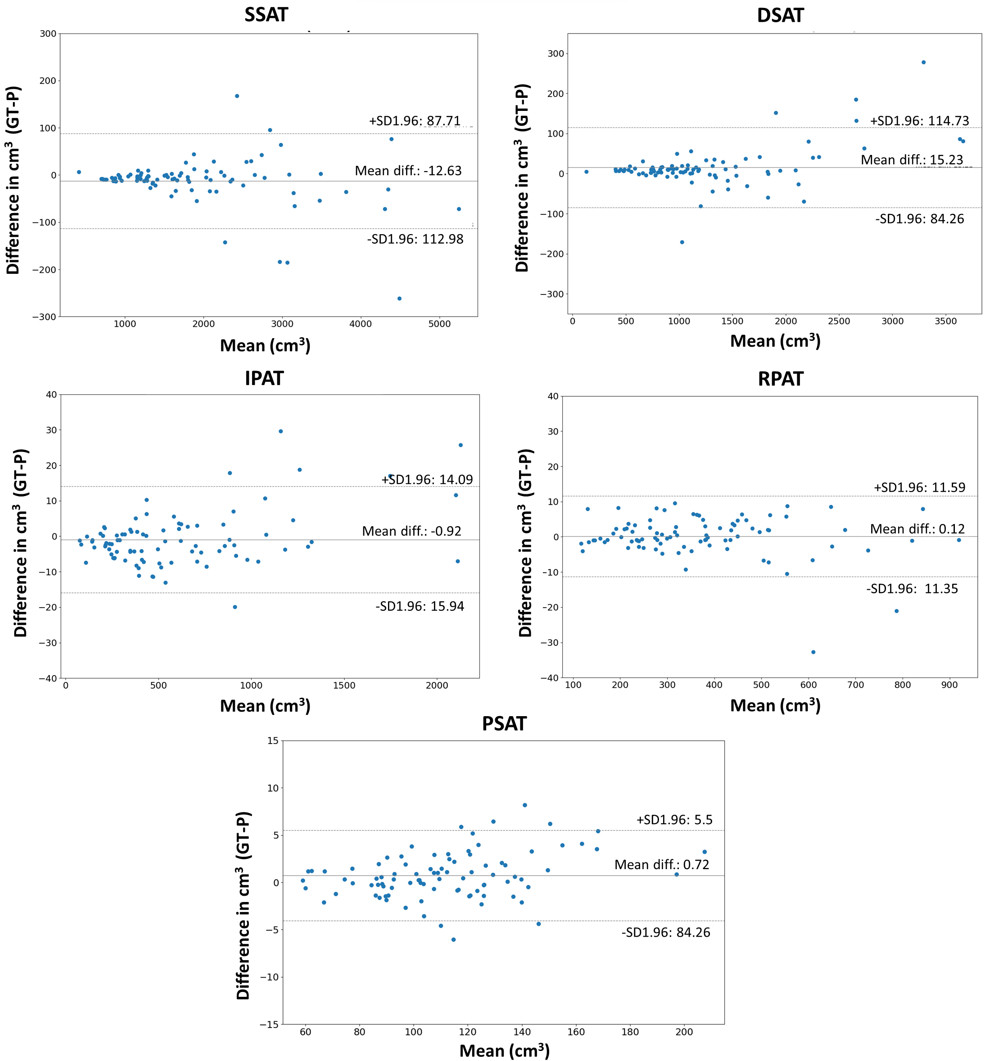
### *Volumetric Performance - Hold-out Test Set*

Figure 2 shows the Bland-Altman plots for AAT predictions on the hold-out test set. The plots indicate robust volumetric assessment for IAT depots with marginal overestimation for IPAT (mean difference: 0.9 ± 7.7 cm3), and marginal underestimation for RPAT (mean difference: 0.1 ± 5.9 cm3) and PSAT (mean difference: 0.7 ± 2.4 cm3). When these under- and over-estimations were expressed as a percentage of the median of the ground truth volumes, the mismatches accounted for only 0.21%, 0.04%, and 0.66% of IPAT, RPAT, and PSAT volumes, respectively. Assessment of the model’s prediction performance for SAT depots indicated a slight overestimation for SSAT (mean difference: 12.6 ± 51.2 cm3) and, consequently, slight underestimation for DSAT (mean difference: 15.2 ± 50.8 cm3). The percentage differences with respect to the median of the ground truth volumes were 0.78% and 1.4% for SSAT and DSAT, respectively. Across all AAT depots, besides 15 unique data points, all estimates were found within the 95% limits of agreement.

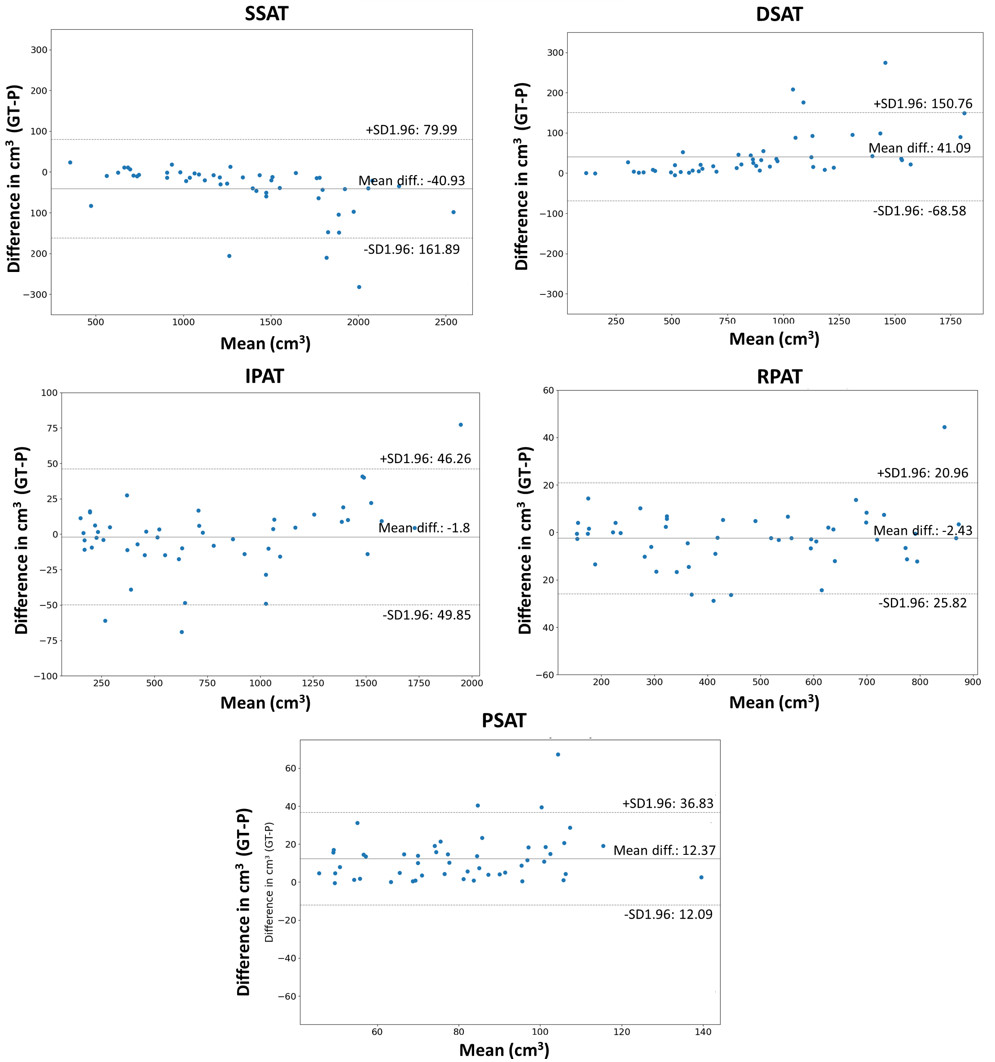
### *Volumetric Performance - External Test Set*

Figure 3 shows the volumetric agreement between predictions from model 1 and the ground truth volumes of the external test set. The model predictions showed marginal overestimation for IPAT (mean difference: 1.8 ± 24.8 cm3) and RPAT (mean difference: 2.4 ± 12.1 cm3). With respect to the median of the depot-specific ground truth volumes, the mean differences were as low as 0.3% and 0.6% for IPAT and RPAT, respectively. Estimates for DSAT (mean difference: 41.1 ± 56.5 cm3) and PSAT (mean difference: 12.4 ± 12.6 cm3) showed a tendency for underestimation, while estimates for SSAT (mean difference: 40.9 ± 62.3 cm3) showed a slight tendency for overestimation. The volumetric differences, expressed as a percentage of the ground truth volumes, were 4.7%, 14.7%, and 3.1% for DSAT, PSAT, and SSAT, respectively. In summary, while the mean estimation errors accounted for less than 5% of the individual AAT depots, SSAT, DSAT, IPAT, and RPAT, increased volumetric difference was observed for PSAT estimates (average volumetric deviation of 14.7%). When comparing the segmentation performance between the internal hold-out test set and the external test set, we observed a slight reduction in volumetric accuracy, indicated in Dice and other quantitative evaluation metrics (Table 2). Although, the model’s application on the external test set showed an overall slight drop in performance, only 6 unique data points were identified as outliers and all the remaining estimates were found within the 95% limits of agreement.

**Figure 2. Bland-Altman plots for predictions on the hold-out test set (S-PRESTO cohort).** Volumetric agreement between predicted and ground truth volumes. Volumetric difference in cubic centimetres (cm3) between ground truth (GT) volumes and predicted (P) volumes from model 1, on the hold-out test set. SSAT = Superficial Subcutaneous Adipose Tissue, DSAT = Deep Subcutaneous Adipose Tissue, IPAT = Intraperitoneal Adipose Tissue, RPAT = Retroperitoneal Adipose Tissue, PSAT = Paraspinal Adipose Tissue.



**Figure 3. Bland-Altman plots for predictions on the external test set (SAMS cohort).** Volumetric agreement between predicted and ground truth volumes. Volumetric difference in cubic centimetres (cm3) between ground truth (GT) volumes and predicted (P) volumes from model 1, on the external test set. SSAT = Superficial Subcutaneous Adipose Tissue, DSAT = Deep Subcutaneous Adipose Tissue, IPAT = Intraperitoneal Adipose Tissue, RPAT = Retroperitoneal Adipose Tissue, PSAT = Paraspinal Adipose Tissue.



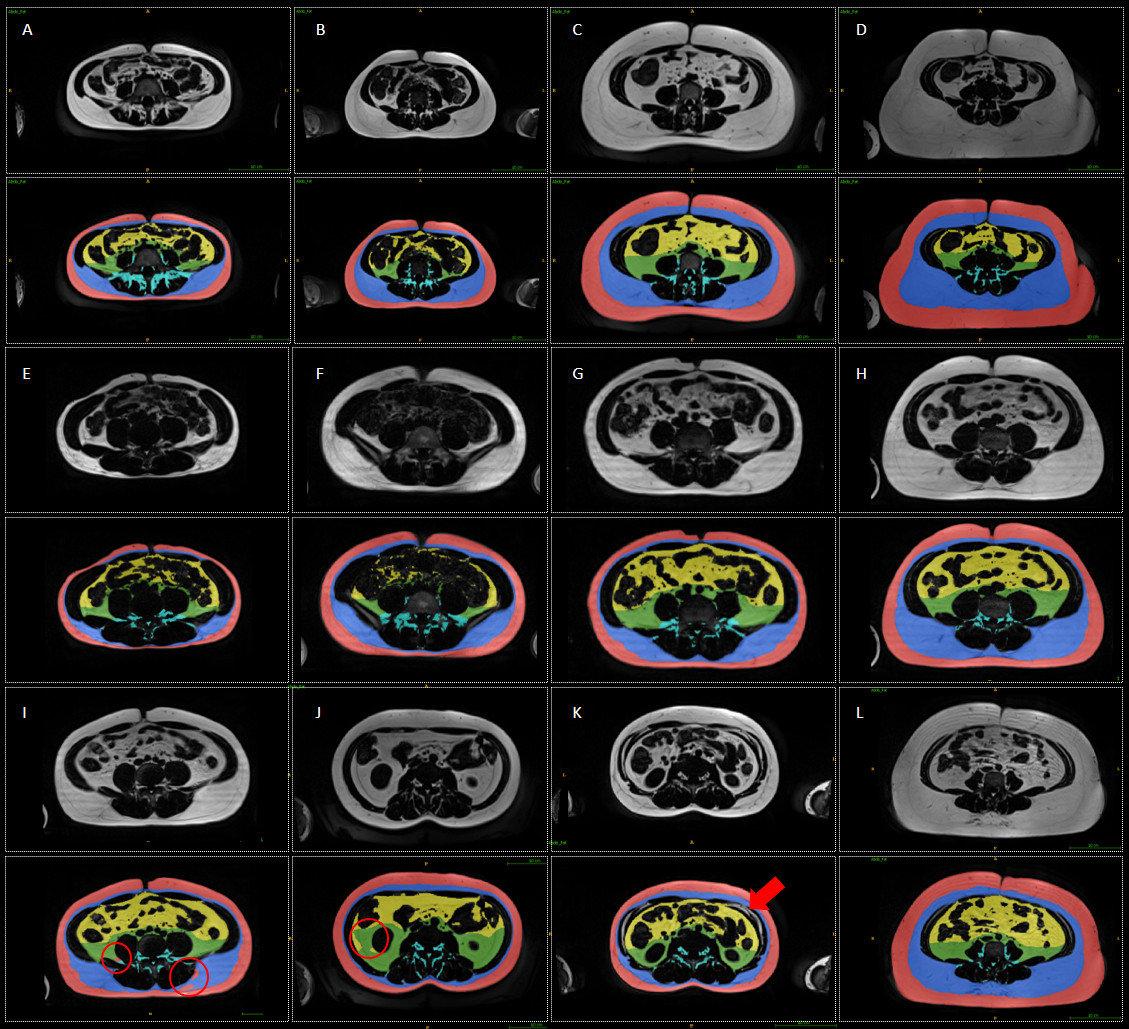
# Discussion

## *Key findings*

A FCNN was implemented to determine whether a deep learning model can be used to accurately quantify individual AAT sub-depots automatically. While previous work has shown successful implementations of FCNNs to automatically quantify total AAT, SAT (including SSAT and DSAT), and VAT in medical imaging data (Kway, Y.M., 2021; Wang et al., 2017; Weston et al., 2019; Dabiri et al., 2020; Estrada et al., 2020; Küstner et al., 2020), the implementation of a FCNN to quantify AAT sub-depots IPAT, RPAT, and PSAT has not been explored. To the best of our knowledge, this is the first study to implement a 3D deep learning method to segment AAT sub-depots, SSAT, DSAT, IPAT, RPAT, and PSAT. Because all these sub-depots express differences in their biological functions and associations with obesity-related complications (Hadar et al., 1983; Arner, 1998; Item and Konrad, 2012; Lee et al., 2013; Chen et al., 2014; Kwok et al., 2016), accurate differentiation is important to better understand the complex manifestation of obesity.

The presented FCNN was evaluated using 5-fold cross-validation. The proposed method showed high accuracy when compared with manually created ground truth data. We found slightly increased segmentation accuracy for SAT depots (Dice, SSAT = 98.21%, DSAT = 97.29%) compared to IAT depots (Dice, IPAT= 96.76%, RPAT= 96.50%, and PSAT=96.08%; all P<0.05). This difference is most likely attributed to the disparity in distributional patterns and anatomical location between SAT and IAT depots. SAT sub-depots are single and continuous depots and are located directly below the epidermis. IAT depots on the other hand, are discontinuous, spatially scattered within the abdominal cavity, and do not follow a fixed structure or form, which makes depot specific classification of IAT more challenging. Furthermore, DSAT can be separated from SSAT by the fascia superficialis, while IAT depots cannot be separated by a clear and traceable boundary line and need to be distinguished based on defined anatomical spaces (Figure S1 [supplement]). Despite the above-mentioned challenges, the proposed model segments distinct IAT depots with a high level of accuracy (Dice>96%).

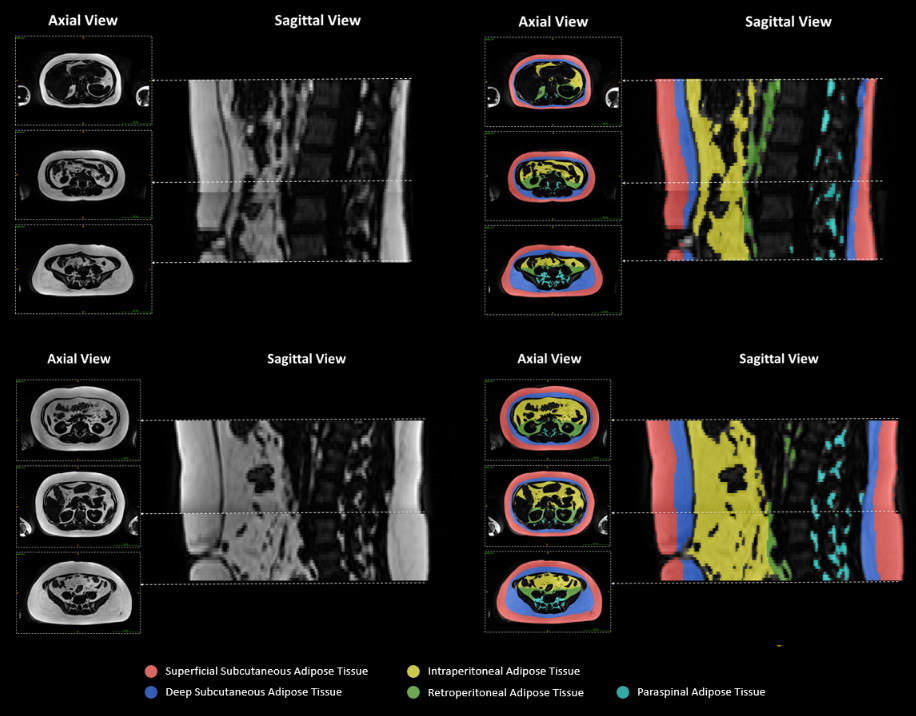
**Figure 4. Representative segmentations of model 1 prediction on the hold-out test set (S-PRESTO) and external test set (SAMS).** Superficial subcutaneous adipose tissue (red), deep subcutaneous adipose tissue (blue), intraperitoneal adipose tissue (yellow), retroperitoneal adipose tissue (green), and paraspinal adipose tissue (cyan). Examples for robust segmentation performance over a wide range of different weight categories for individuals from the hold-out test set (A-D) and the external test set (E-H). Misclassification examples for prediction on the external test set (I) and hold-out test set (J). (K) Shows the strength of the algorithm – excluding intramuscular adipose tissue (adipose tissue between the abdominal cavity and abdominal wall). (L) Illustrates robustness of the model to ghosting artifacts (ghosting of the fascia superficialis, most prominent in the anterior portion of the abdomen).



Bland-Altman plots showed a high accordance between the predicted and ground truth volumes, on the hold-out (Figure 2) and the external test set (Figure 3). The plots indicated robust segmentation performance across a wide range of weight categories, from individuals who are underweight to individuals affected by overweight and obesity, with almost all estimates falling within the 95% limits of agreement. In addition to Bland-Altman plots, visualization of the predicted segmentation masks (Figure 4, hold-out test set (A-D), external test set (E-H)) affirmed robust segmentation performance across a wide range of depo-specific fat accumulation (min-max: SSAT (402.7 - 5480.5 cm3), DSAT (134.3 - 4759.7 cm3), IPAT (53.3 - 2139.4 cm3), RPAT (85.4 - 1140.5 cm3), and PSAT (48.6 - 209.0 cm3)). Collectively, our results indicate that the presented method is capable of producing accurate segmentation across individuals with diverse depot-specific AAT accumulation and weight levels. Therefore, the presented method could offer robust and rapid AAT quantification for population-based data. The presented model generalized well from training domain, hold-out test set (S-PRESTO, female cohort), to external test set (SAMS, male cohort), which indicated robust generalization to a different scanner, when using Dixon sequences. Only a slight reduction in segmentation performance was observed for SSAT (Dice, 98.2% (hold-out) 🡪 95.7% (external test)), DSAT (Dice, 97.3% (hold-out) 🡪 94.5% (external test)), IPAT (Dice, 96.8% (hold-out) 🡪 93.5% (external test)), and RPAT (Dice, 96.5% (hold-out) 🡪 92.7% (external test)), while higher performance drop was observed for PSAT (Dice, 96.1% (hold-out) 🡪 88.5% (external test)). The performance differences observed for PSAT quantification between the external test set (comprised of men) and the training data set (exclusively women) might be contributed to morphological sexual dimorphism in the lumbar spine (Hay et al., 2015). Further investigations and analyses are warranted to explore this hypothesis and gain a deeper understanding of the factors influencing the performance differences observed. Interestingly, the performance difference between the hold-out test set and the external test set was characterized by a notable increase in the standard deviation for the HD95 metric. This higher standard deviation in HD95 in the external test set might suggest the presence of elevated localized prediction discrepancies, such as small clusters of misclassified pixels. However, despite these localized discrepancies, the high Dice scores (>90%) indicate a preserved overall overlap agreement between the predictions and the ground truth data of the external test set. A potential example that could illustrate this phenomenon is shown in Figure 4, I.

Detailed visual examination of the predicted segmentation masks indicated consistent performance over all levels of the abdomen (Figure 5). The visualization also revealed minor misclassifications between RPAT and IPAT, in the hold-out test set (Figure 4, J) and between SSAT and other adipose depots in the external test set (Figure 4, I). However, volumetric inspections revealed that only a few image slices were affected by these misclassifications and, therefore, these errors did not contribute significantly to the absolute volumetric quantification of the respective AAT depots, as highlighted by Bland-Altman plots (Figure 2 and 3) and quantitative evaluation metrics (Table 2). The visualization of the predicted segmentation masks also revealed certain strengths of the proposed algorithm. Given the high-quality training data, the model successfully learned to exclude adipose tissue located between the abdominal wall and the abdominal cavity (intramuscular adipose tissue which does not classify as PSAT). A visual example can be found in Figure 4 (K). Additionally, it was observed that reliable segmentation of SSAT and DSAT is still preserved in the presence of artifacts, which affect the appearance of the fascia superficialis (Figure 4 (L)).

**Figure 5. Visual segmentation example over different levels of the abdomen.** Superficial subcutaneous adipose tissue (red), deep subcutaneous adipose tissue (blue), intraperitoneal adipose tissue (yellow), retroperitoneal adipose tissue (green), and paraspinal adipose tissue (cyan). This figure illustrates the segmentation performance over different levels of the abdomen for a normal weight (BMI<23, top) and overweight participant (BMI>23, bottom).



## *Strengths and Limitations*

The strengths of the study include comprehensive segmentation and quantification of SAT into SSAT and DSAT, and IAT into IPAT, RPAT, and PSAT, enabling detailed assessment of abdominal obesity. To facilitate this study, high-quality datasets were meticulously created, consisting of a total of 439 3D MRI volumes that were acquired across two different populations using different scanners. The data included both men and women and covered a wide range of weight categories. This diverse sample allowed for a comprehensive representation of distinct anatomical AAT depots. While our model demonstrated robust generalization capabilities on Dixon sequences within the studied population, it is important to acknowledge that the validation results are limited to the available data. Therefore, it is crucial to conduct further investigations to assess the model's ability to generalize to more diverse datasets, including populations such as children and individuals with morbid obesity, as well as data obtained using different MRI protocols (variations in field strength and imaging sequence).

# Conclusions

The proposed model offers accurate quantification of AAT sub-depots, thereby enabling a more detailed characterization of abdominal obesity. Imaging studies typically quantify AAT in a cross-sectional image slice due to cost and time constraints. However, cross-sectional images might not appropriately account for ethnic, sex, and age differences (Demerath et al., 2007), and the position of the cross-section can influence the magnitude of the association with clinical measurements (Lee et al., 2011; Brown et al., 2015). The proposed method can address these challenges by providing fast, automated, and standardized volumetric quantification of AAT depots. Therefore, the proposed method paves the way for population-based and longitudinal studies to identify different obesity phenotypes and opens opportunities to study metabolic risks linked to individual AAT sub-depots. Particularly, the study of adipose tissue biology has been limited to specific sub-populations and small groups, such as individuals undergoing bariatric surgery, in which tissue sampling of distinct AAT depots is feasible. When combined with advanced imaging methods to assess adipose tissue lipid content and fatty acid composition (Lunati et al., 2001; Hamilton et al., 2017), the present method could contribute to advancements in current understanding of AAT depot-specific biology. While further examination of the model’s generalization to other imaging protocols and populations is needed in future work, the proposed work can aid in the understanding of the complex manifestations of obesity and its concomitant medical complications; hence, it might contribute to improvements in the assessment and treatment of obesity and cardiometabolic disease.

Considering our study for future research, two primary directions emerge. To enhance our model's robustness and versatility, it's essential to broaden the evaluation and application of the proposed model to a wider range of data, including paediatric and elderly populations, and on individuals that are affected by lipodystrophies. Extending the application of the proposed model to these diverse population groups holds substantial promise for advancing our understanding of adipose tissue distribution and its variations across different demographic categories and disease conditions. Another promising avenue for future research involves further refining our model to allow for an even more granular assessment of AAT. While the presented model already provides a more extensive assessment when compared to existing methodologies, future endeavours could delve deeper into distinguishing even more specific adipose tissue depots, such as mesenteric, omental, and perirenal adipose tissue. While future directions can expand upon the foundation laid by our current study, enhancing the versatility and precision of the proposed model, in conclusion, our work represents a significant stride in advancing precision imaging for obesity.

**Code availability**

The modelling code is publicly available at GitHub: <https://github.com/YesheKway/AAT_seg>.

**Funding**

The study is supported by the National Research Foundation (NRF) under the Open Fund-Large Collaborative Grant (OF-LCG; MOH-000504) administered by the Singapore Ministry of Health’s National Medical Research Council (NMRC) and the Agency for Science, Technology and Research (A\*STAR). In RIE2025, S-PRESTO is supported by funding from the NRF’s Human Health and Potential (HHP) Domain, under the Human Potential Programme.

K.M.G. is supported by the UK Medical Research Council (MC\_UU\_12011/4), the National Institute for Health Research (NIHR Senior Investigator (NF-SI-0515-10042) and NIHR Southampton Biomedical Research Centre (NIHR203319)) and the US National Institute on Aging of the National Institutes of Health (Award No. U24AG047867). For the purpose of Open Access, the author has applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising from this submission. M.F. is supported by the National Research Foundation Singapore under its AI Singapore Programme (Award Number: AISG-GC-2019-001-2A) and the NMRC Health Service Research Grant (MOH-000030-00).

The funders had no role in the data collection and analysis, and the  
decision to submit for publication.

**Conflicts of interest**

K.M.G. has received reimbursement for speaking at conferences sponsored by companies selling nutritional products. K.M.G., S-Y.C., and Y.S.C. are part of an academic consortium that has received research funding from Abbott Nutrition, Nestec, BenevolentAI Bio Ltd., and Danone. Other authors declare no conflict of interest.

**Acknowledgements**

We thank the nonauthor members of this study group: Airu Chia, Andrea Cremaschi, Anna Magdalena Fogel, Anne Eng Neo Goh, Anne Rifkin-Graboi, Anqi Qiu, Bee Wah Lee, Bernard Su Min Chern, Candida Vaz, Chan Shi Yu, Dawn Xin Ping Koh, Dennis Wang, Desiree Y. Phua, Elaine Phaik Ling Quah, Elizabeth Huiwen Tham, Evelyn Chung Ning Law, Evelyn Keet Wai Lau, Evelyn Xiu Ling Loo, Fabian Kok Peng Yap, Falk Müller-Riemenschneider, George Seow Heong Yeo, Gerard Chung Siew Keong, Hannah Ee Juen Yong, Helen Yu Chen, Hong Pan, Huang Jian, Huang Pei, Hugo P S van Bever, Hui Min Tan, Ives Lim Yubin, Jadegoud Yaligar, Jia Xu, Jonathan Tze Liang Choo, Jonathan Y. Bernard, Jonathan Yinhao Huang, Jun Shi Lai, Karen Mei Ling Tan, Keri McCrickerd, Kok Wee Chong, Kothandaraman Narasimhan, Kuan Jin Lee, Li Chen, Lieng Hsi Ling, Ling-Wei Chen, Lourdes Mary Daniel, Lynette Pei-Chi Shek, Maria De Iorio, Mary Foong-Fong Chong, Mary Wlodek, Mei Chien Chua, Michael J. Meaney, Michelle Zhi Ling Kee, Min Gong, Mya Thway Tint, Neerja Karnani, Ngee Lek, Noor Hidayatul Aini Bte Suaini, Oon Hoe Teoh, Priti Mishra, Queenie Ling Jun Li, See Ling Loy, Seng Bin Ang, Shirong Cai, Shu-E Soh, Si Hui Goh, Stephen Chin-Ying Hsu, Tan Ai Peng, Teng Hong Tan, Varsha Gupta, Victor Samuel Rajadurai, Wee Meng Han, Wei Wei Pang, Yin Bun Cheung, Yiong Huak Chan, Zai Ru Cheng, Zhang Han

**References**

Abate, N., Burns, D., Peshock, R.M., Garg, A., Grundy, S.M., 1994. Estimation of adipose tissue mass by magnetic resonance imaging: validation against dissection in human cadavers. Journal of lipid research 35 (8), 1490–1496.

Anoop, S., Misra, A., Bhatt, S.P., Gulati, S., Mahajan, H., Prabakaran, G., 2017. High Plasma Glucagon Levels Correlate with Waist-to-Hip Ratio, Suprailiac Skinfold Thickness, and Deep Subcutaneous Abdominal and Intraperitoneal Adipose Tissue Depots in Nonobese Asian Indian Males with Type 2 Diabetes in North India. Journal of diabetes research 2017, 2376016. 10.1155/2017/2376016.

Arner, P., 1998. Not all fat is alike. The Lancet 351 (9112), 1301–1302. 10.1016/S0140-6736(05)79052-8.

Bhanu, P.K., Arvind, C.S., Yeow, L.Y., Chen, W.X., Lim, W.S., Tan, C.H., 2021. CAFT: a deep learning-based comprehensive abdominal fat analysis tool for large cohort studies. Magma (New York, N.Y.). 10.1007/s10334-021-00946-9.

Borga, M., 2018. MRI adipose tissue and muscle composition analysis-a review of automation techniques. The British journal of radiology 91 (1089), 20180252. 10.1259/bjr.20180252.

Brown, R.E., Kuk, J.L., Lee, S., 2015. Measurement site influences abdominal subcutaneous and visceral adipose tissue in obese adolescents before and after exercise. Pediatric obesity 10 (2), 98–104. 10.1111/j.2047-6310.2014.224.x.

Chen, Y.-Y., Pao, J.-L., Liaw, C.-K., Hsu, W.-L., Yang, R.-S., 2014. Image changes of paraspinal muscles and clinical correlations in patients with unilateral lumbar spinal stenosis. European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society 23 (5), 999–1006. 10.1007/s00586-013-3148-z.

Dabiri, S., Popuri, K., Ma, C., Chow, V., Feliciano, E.M.C., Caan, B.J., Baracos, V.E., Beg, M.F., 2020. Deep learning method for localization and segmentation of abdominal CT. Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society 85, 101776. 10.1016/j.compmedimag.2020.101776.

Demerath, E.W., Sun, S.S., Rogers, N., Lee, M., Reed, D., Choh, A.C., Couch, W., Czerwinski, S.A., Chumlea, W.C., Siervogel, R.M., Towne, B., 2007. Anatomical patterning of visceral adipose tissue: race, sex, and age variation. Obesity (Silver Spring, Md.) 15 (12), 2984–2993. 10.1038/oby.2007.356.

Estrada, S., Lu, R., Conjeti, S., Orozco-Ruiz, X., Panos-Willuhn, J., Breteler, M.M.B., Reuter, M., 2020. FatSegNet: A fully automated deep learning pipeline for adipose tissue segmentation on abdominal dixon MRI. Magnetic resonance in medicine 83 (4), 1471–1483. 10.1002/mrm.28022.

Freedland, E.S., 2004. Role of a critical visceral adipose tissue threshold (CVATT) in metabolic syndrome: implications for controlling dietary carbohydrates: a review. Nutrition & metabolism 1 (1), 12. 10.1186/1743-7075-1-12.

Gantz, M., Chen, J., B. Heymsfield, S., Shen, W., 2011. Intraperitoneal Adipose Tissue: Associated Health Risks, Quantification by Advanced Imaging Methods and Future Directions in Children. The Open Obesity Journal 3 (1).

Hadar, H., Gadoth, N., Heifetz, M., 1983. Fatty replacement of lower paraspinal muscles: normal and neuromuscular disorders. AJR. American journal of roentgenology 141 (5), 895–898. 10.2214/ajr.141.5.895.

Hamilton, G., Schlein, A.N., Middleton, M.S., Hooker, C.A., Wolfson, T., Gamst, A.C., Loomba, R., Sirlin, C.B., 2017. In vivo triglyceride composition of abdominal adipose tissue measured by 1 H MRS at 3T. Journal of magnetic resonance imaging : JMRI 45 (5), 1455–1463. 10.1002/jmri.25453.

Hay, O., Dar, G., Abbas, J., Stein, D., May, H., Masharawi, Y., Peled, N., Hershkovitz, I., 2015. The Lumbar Lordosis in Males and Females, Revisited. PloS one 10 (8), e0133685. 10.1371/journal.pone.0133685.He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep Residual Learning for Image Recognition. <https://arxiv.org/pdf/1512.03385>.

Hu, H.H., Chen, J., Shen, W., 2016. Segmentation and quantification of adipose tissue by magnetic resonance imaging. Magma (New York, N.Y.) 29 (2), 259–276. 10.1007/s10334-015-0498-z.

Ibrahim, M.M., 2010. Subcutaneous and visceral adipose tissue: structural and functional differences. Obesity reviews : an official journal of the International Association for the Study of Obesity 11 (1), 11–18. 10.1111/j.1467-789X.2009.00623.x.

Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P.F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., Wirkert, S., Maier-Hein, K.H., 2018. nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation. <https://arxiv.org/pdf/1809.10486>.

Item, F., Konrad, D., 2012. Visceral fat and metabolic inflammation: the portal theory revisited. Obesity reviews : an official journal of the International Association for the Study of Obesity 13 Suppl 2, 30–39. 10.1111/j.1467-789X.2012.01035.x.

Kalichman, L., Carmeli, E., Been, E., 2017. The Association between Imaging Parameters of the Paraspinal Muscles, Spinal Degeneration, and Low Back Pain. BioMed research international 2017, 2562957. 10.1155/2017/2562957.

Katzmarzyk, P.T., Shen, W., Baxter-Jones, A., Bell, J.D., Butte, N.F., Demerath, E.W., Gilsanz, V., Goran, M.I., Hirschler, V., Hu, H.H., Maffeis, C., Malina, R.M., Müller, M.J., Pietrobelli, A., Wells, J.C.K., 2012. Adiposity in children and adolescents: correlates and clinical consequences of fat stored in specific body depots. Pediatric obesity 7 (5), e42-61. 10.1111/j.2047-6310.2012.00073.x.

Kelley, D.E., Thaete, F.L., Troost, F., Huwe, T., Goodpaster, B.H., 2000. Subdivisions of subcutaneous abdominal adipose tissue and insulin resistance. American journal of physiology. Endocrinology and metabolism 278 (5), E941-8. 10.1152/ajpendo.2000.278.5.E941.

Khoo, C.M., Leow, M.K.-S., Sadananthan, S.A., Lim, R., Venkataraman, K., Khoo, E.Y.H., Velan, S.S., Ong, Y.T., Kambadur, R., McFarlane, C., Gluckman, P.D., Lee, Y.S., Chong, Y.S., Tai, E.S., 2014. Body fat partitioning does not explain the interethnic variation in insulin sensitivity among Asian ethnicity: the Singapore adults metabolism study. Diabetes 63 (3), 1093–1102. 10.2337/db13-1483.

Kingma, D.P., Ba, J., 2014. Adam: A Method for Stochastic Optimization. <https://arxiv.org/pdf/1412.6980>.

Kullberg, J., Hedström, A., Brandberg, J., Strand, R., Johansson, L., Bergström, G., Ahlström, H., 2017. Automated analysis of liver fat, muscle and adipose tissue distribution from CT suitable for large-scale studies. In Scientific reports 7 (1), p. 10425. DOI: 10.1038/s41598-017-08925-8.

Küstner, T., Hepp, T., Fischer, M., Schwartz, M., Fritsche, A., Häring, H.-U., Nikolaou, K., Bamberg, F., Yang, B., Schick, F., Gatidis, S., Machann, J., 2020. Fully Automated and Standardized Segmentation of Adipose Tissue Compartments by Deep Learning in Three-dimensional Whole-body MRI of Epidemiological Cohort Studies. <https://arxiv.org/pdf/2008.02251>.

Kway, Y.M., Thirumurugan, K., Tint, M.T., Michael, N., Shek, L.P.-C., Yap, F.K.P., Tan, K.H., Godfrey, K.M., Chong, Y.S., Fortier, M.V., Marx, U.C., Eriksson, J.G., Lee, Y.S., Velan, S.S., Feng, M., Sadananthan, S.A., 2021. Automated Segmentation of Visceral, Deep Subcutaneous, and Superficial Subcutaneous Adipose Tissue Volumes in MRI of Neonates and Young Children. Radiology: Artificial Intelligence 3 (5), e200304. 10.1148/ryai.2021200304.

Kwok, K.H.M., Lam, K.S.L., Xu, A., 2016. Heterogeneity of white adipose tissue: molecular basis and clinical implications. Exp Mol Med 48 (3), e215. 10.1038/emm.2016.5.

Lee, M.-J., Wu, Y., Fried, S.K., 2013. Adipose tissue heterogeneity: implication of depot differences in adipose tissue for obesity complications. Molecular aspects of medicine 34 (1), 1–11. 10.1016/j.mam.2012.10.001.

Lee, S., Kuk, J.L., Kim, Y., Arslanian, S.A., 2011. Measurement site of visceral adipose tissue and prediction of metabolic syndrome in youth. Pediatric diabetes 12 (3 Pt 2), 250–257. 10.1111/j.1399-5448.2010.00705.x.

Leitner, B.P., Huang, S., Brychta, R.J., Duckworth, C.J., Baskin, A.S., McGehee, S., Tal, I., Dieckmann, W., Gupta, G., Kolodny, G.M., Pacak, K., Herscovitch, P., Cypess, A.M., Chen, K.Y., 2017. Mapping of human brown adipose tissue in lean and obese young men. Proceedings of the National Academy of Sciences of the United States of America 114 (32), 8649–8654. 10.1073/pnas.1705287114.

Long, J., Shelhamer, E., Darrell, T., 2014. Fully Convolutional Networks for Semantic Segmentation. <https://arxiv.org/pdf/1411.4038>.

Loo, E.X.L., Soh, S.-E., Loy, S.L., Ng, S., Tint, M.T., Chan, S.-Y., Huang, J.Y., Yap, F., Tan, K.H., Chern, B.S.M., Tan, H.H., Meaney, M.J., Karnani, N., Godfrey, K.M., Lee, Y.S., Chan, J.K.Y., Gluckman, P.D., Chong, Y.-S., Shek, L.P.-C., Eriksson, J.G., Chia, A., Fogel, A.M., Goh, A.E.N., Chu, A.H.Y., Rifkin-Graboi, A., Qiu, A., Lee, B.W., Cheon, B.K., Vaz, C., Henry, C.J., Forde, C.G., Chi, C., Koh, D.X.P., Phua, D.Y., Loh, D.N.L., Quah, E.P.L., Tham, E.H., Law, E.C.N., Magkos, F., Mueller-Riemenschneider, F., Yeo, G.S.H., Yong, H.E.J., Chen, H.Y., Pan, H., van Bever, H.P.S., Tan, H.M., Aris, I.B.M., Tay, J., Xu, J., Yoong, J.S.-Y., Eriksson, J.G., Choo, J.T.L., Bernard, J.Y., Lai, J.S., Tan, K.M.L., Kwek, K.Y.C., McCrickerd, K., Narasimhan, K., Chong, K.W., Lee, K.J., Chen, L., Ling, L.H., Chen, L.-W., Daniel, L.M., Fortier, M.V., Chong, M.F.-F., Chua, M.C., Leow, M.K.-S., Kee, M.Z.L., Gong, M., Michael, N., Lek, N., Teoh, O.H., Mishra, P., Li, Q.L.J., Velan, S.S., Ang, S.B., Cai, S., Goh, S.H., Lim, S.B., Tsotsi, S., Hsu, S.C.-Y., Toh, S.-A.E.S., Sadananthan, S.A., Tan, T.H., Yew, T.W., Gupta, V., Rajadurai, V.S., Han, W.M., Pang, W.W., Yuan, W.L., Zhu, Y., Cheung, Y.B., Chan, Y.H., Cheng, Z.R., 2021. Cohort profile: Singapore Preconception Study of Long-Term Maternal and Child Outcomes (S-PRESTO). European journal of epidemiology 36 (1), 129–142. 10.1007/s10654-020-00697-2.

Lunati, E., Marzola, P., Nicolato, E., Sbarbati, A., 2001. In-vivo quantitative hydrolipidic map of perirenal adipose tissue by chemical shift imaging at 4.7 Tesla. International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity 25 (4), 457–461. 10.1038/sj.ijo.0801262.

Lundbom, J., Hakkarainen, A., Lundbom, N., Taskinen, M.-R., 2013. Deep subcutaneous adipose tissue is more saturated than superficial subcutaneous adipose tissue. International journal of obesity (2005) 37 (4), 620–622. 10.1038/ijo.2012.72.

Marinou, K., Hodson, L., Vasan, S.K., Fielding, B.A., Banerjee, R., Brismar, K., Koutsilieris, M., Clark, A., Neville, M.J., Karpe, F., 2014. Structural and functional properties of deep abdominal subcutaneous adipose tissue explain its association with insulin resistance and cardiovascular risk in men. Diabetes care 37 (3), 821–829. 10.2337/dc13-1353.

Miyazaki, Y., Glass, L., Triplitt, C., Wajcberg, E., Mandarino, L.J., DeFronzo, R.A., 2002. Abdominal fat distribution and peripheral and hepatic insulin resistance in type 2 diabetes mellitus. American journal of physiology. Endocrinology and metabolism 283 (6), E1135-43. 10.1152/ajpendo.0327.2001.

Pérez-García, F., Sparks, R., Ourselin, S., 2021. TorchIO: A Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. Computer Methods and Programs in Biomedicine 208, 106236. 10.1016/j.cmpb.2021.106236.

Richard Shaw, Carole Sudre, Sebastien Ourselin, M. Jorge Cardoso, 2019. MRI k-Space Motion Artefact Augmentation: Model Robustness and Task-Specific Uncertainty. International Conference on Medical Imaging with Deep Learning, 427–436.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), Medical image computing and computer-assisted intervention - MICCAI 2015. 18th International Conference, Munich, Germany, October 5-9, 2015, proceedings / Nassir Navab, Joachim Hornegger, William M. Wells, Alejandro F. Frangi (eds.), vol. 9351. Springer, Cham, pp. 234–241.

Sadananthan, S.A., Tint, M.T., Michael, N., Aris, I.M., Loy, S.L., Lee, K.J., Shek, L.P.-C., Yap, F.K.P., Tan, K.H., Godfrey, K.M., Leow, M.K.-S., Lee, Y.S., Kramer, M.S., Gluckman, P.D., Chong, Y.S., Karnani, N., Henry, C.J., Fortier, M.V., Velan, S.S., 2019. Association Between Early Life Weight Gain and Abdominal Fat Partitioning at 4.5 Years is Sex, Ethnicity, and Age Dependent. Obesity (Silver Spring, Md.) 27 (3), 470–478. 10.1002/oby.22408.

Shen, W., Wang, Z., Punyanita, M., Lei, J., Sinav, A., Kral, J.G., Imielinska, C., Ross, R., Heymsfield, S.B., 2003. Adipose tissue quantification by imaging methods: a proposed classification. Obesity research 11 (1), 5–16. 10.1038/oby.2003.3.

Shuster, A., Patlas, M., Pinthus, J.H., Mourtzakis, M., 2012. The clinical importance of visceral adiposity: a critical review of methods for visceral adipose tissue analysis. The British journal of radiology 85 (1009), 1–10. 10.1259/bjr/38447238.

Tanaka, M., Okada, H., Hashimoto, Y., Kumagai, M., Nishimura, H., Fukui, M., 2020. Intraperitoneal, but not retroperitoneal, visceral adipose tissue is associated with diabetes mellitus: a cross-sectional, retrospective pilot analysis. Diabetology & metabolic syndrome 12 (1), 103. 10.1186/s13098-020-00612-5.

Tanaka, M., Okada, H., Hashimoto, Y., Kumagai, M., Nishimura, H., Fukui, M., 2021. Distinct associations of intraperitoneal and retroperitoneal visceral adipose tissues with metabolic syndrome and its components. Clinical nutrition (Edinburgh, Scotland) 40 (5), 3479–3484. 10.1016/j.clnu.2020.11.030.

Tchernof, A., Després, J.-P., 2013. Pathophysiology of human visceral obesity: an update. Physiological reviews 93 (1), 359–404. 10.1152/physrev.00033.2011.

van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P., 1999. Automated model-based tissue classification of MR images of the brain. IEEE transactions on medical imaging 18 (10), 897–908. 10.1109/42.811270.

Villarroya, F., Cereijo, R., Villarroya, J., Giralt, M., 2017. Brown adipose tissue as a secretory organ. Nat Rev Endocrinol 13 (1), 26–35. 10.1038/nrendo.2016.136.

Walker, G.E., Marzullo, P., Prodam, F., Bona, G., Di Blasio, A.M., 2014. Obesity modifies expression profiles of metabolic markers in superficial and deep subcutaneous abdominal adipose tissue depots. Endocrine 46 (1), 99–106. 10.1007/s12020-013-0040-x.

Wang, Y., Qiu, Y., Thai, T., Moore, K., Liu, H., Zheng, B., 2017. A two-step convolutional neural network based computer-aided detection scheme for automatically segmenting adipose tissue volume depicting on CT images. Computer Methods and Programs in Biomedicine 144, 97–104. 10.1016/j.cmpb.2017.03.017.

Weston, A.D., Korfiatis, P., Kline, T.L., Philbrick, K.A., Kostandy, P., Sakinis, T., Sugimoto, M., Takahashi, N., Erickson, B.J., 2019. Automated Abdominal Segmentation of CT Scans for Body Composition Analysis Using Deep Learning. Radiology 290 (3), 669–679. 10.1148/radiol.2018181432.

Xavier Glorot, Yoshua Bengio, 2010. Understanding the difficulty of training deep feedforward neural networks. undefined.

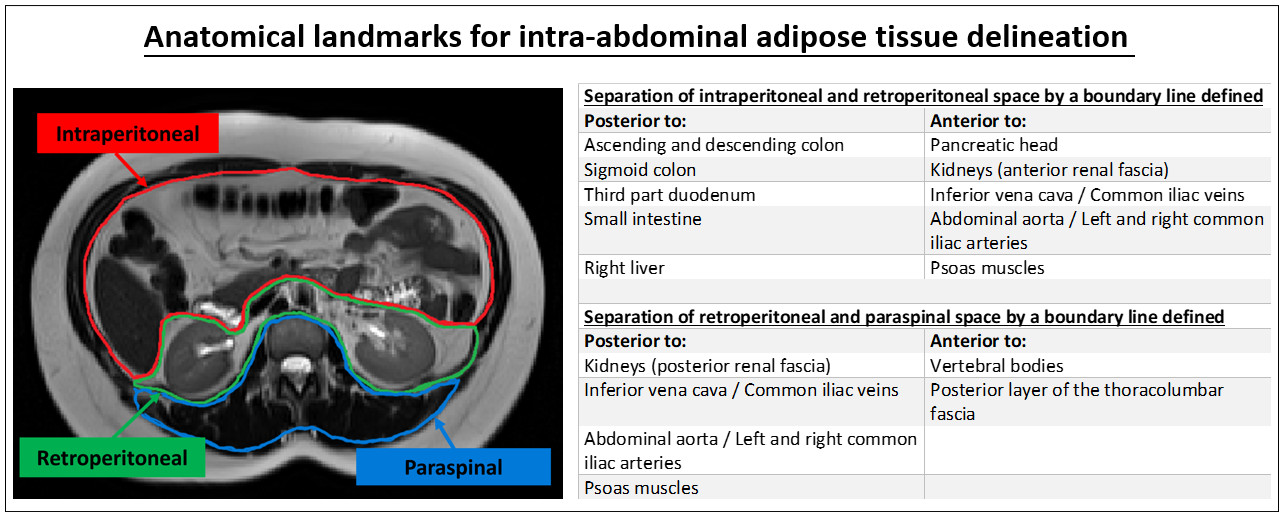
Xu, B., Wang, N., Chen, T., Li, M., 2015. Empirical Evaluation of Rectified Activations in Convolutional Network. <https://arxiv.org/pdf/1505.00853>.

Yang, Y.-K., Chen, M., Clements, R.H., Abrams, G.A., Aprahamian, C.J., Harmon, C.M., 2008. Human mesenteric adipose tissue plays unique role versus subcutaneous and omental fat in obesity related diabetes. Cellular physiology and biochemistry: International Journal of experimental cellular physiology, biochemistry, and pharmacology 22 (5-6), 531–538. 10.1159/000185527.

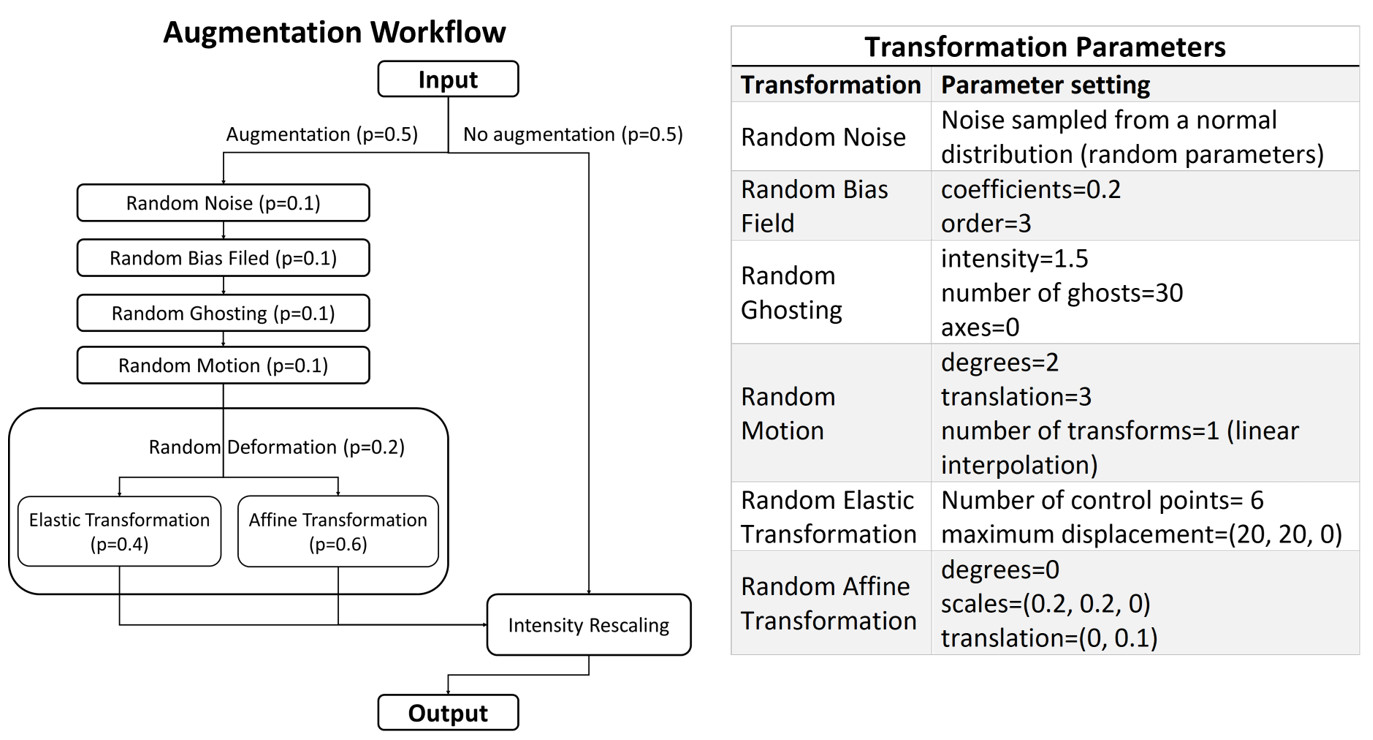
Yanik, B., Keyik, B., Conkbayir, I., 2013. Fatty degeneration of multifidus muscle in patients with chronic low back pain and in asymptomatic volunteers: quantification with chemical shift magnetic resonance imaging. Skeletal radiology 42 (6), 771–778. 10.1007/s00256-012-1545-8.

Zhang, Y., Mandelli, F., Mündermann, A., Nüesch, C., Kovacs, B., Schären, S., Netzer, C., 2021. Association between fatty infiltration of paraspinal muscle, sagittal spinopelvic alignment and stenosis grade in patients with degenerative lumbar spinal stenosis. North American Spine Society Journal 5, 100054. 10.1016/j.xnsj.2021.100054.

**Supplement**



**Figure S1. Anatomical Landmarks for internal abdominal adipose tissue differentiation.** This figure illustrates the segmentation criteria used to differentiate individual internal abdominal adipose tissue depots based on their anatomical location. Adipose tissue voxels below the deep fascia and within the intraperitoneal (red), retroperitoneal (green) and paraspinal (blue) space were labelled as intraperitoneal adipose tissue (IPAT), retroperitoneal adipose tissue (RPAT), and paraspinal adipose tissue (PSAT), respectively. The anatomical spaces were delineated from each other by drawing a boundary line as described.



**Figure S2. Data Augmentation Workflow.** The data augmentation pipeline was implemented using TorchIO. This augmentation pipeline is called on every imaging volume during the training process. Numbers in breakers indicate the probability of an event happening. As indicated in the graph, every time a volume is loaded, there is a 50% change that augmentation is performed. Preliminarily experiments showed that an increase in augmentation probability (>50%) results in decreased segmentation performance on the source domain data set (internal hold out-set). Application of a single augmentation module (e.g., random noise) does not exclude the additional application of other transformations. When random deformation is applied (p=20%), the type of random deformation is either elastic transformation or affine transformation with a probability of 40%, and 60%, respectively. The reason for increased chance of affine transformation is to trade off computational cost/ complexity of deformation and data flow during network training. When imaging volumes are interpolated in the process of any above listed transformation, corresponding label masks are interpolated using nearest neighbour interpolation to match the segmentation mask with the transformed imaging volume. Transformation parameters were selected manually to mimic realistic artifacts, as seen in abdominal MRI. Selected parameters for specific transformation operations are shown in the table (right). Please refer to (https://torchio.readthedocs.io/) for further implementation details.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table S1. 5-fold cross-validation: mean performance over all test sets** | | | | | | | | | | | | |
| **Fat Depot** |  | **Dice**  **(%)** |  | **HD95**  **(mm)** |  | **FP**  **(%)** |  | **FN**  **(%)** |  | **Precision**  **(%)** |  | **Sensitivity**  **(%)** |
| **SSAT** |  | 98.31 ± 0.08 |  | 1.68 ± 0.82 |  | 1.58 ± 0.23 |  | 0.17 ± 0.03 |  | 98.21 ± 0.20 |  | 98.21 ± 0.24 |
| **DSAT** |  | 97.23 ± 0.13 |  | 1.74 ± 1.11 |  | 2.84 ± 0.31 |  | 0.16 ± 0.03 |  | 97.33 ± 0.50 |  | 96.95 ± 0.48 |
| **IPAT** |  | 96.45 ± 0.22 |  | 1.56 ± 0.51 |  | 3.67 ± 0.67 |  | 0.07 ± 0.01 |  | 96.60 ± 0.27 |  | 97.01 ± 0.37 |
| **RPAT** |  | 96.25 ± 0.25 |  | 1.52 ± 0.15 |  | 3.70 ± 0.20 |  | 0.05 ± 0.01 |  | 96.21 ± 0.50 |  | 96.63 ± 0.17 |
| **PSAT** |  | 95.91 ± 0.34 |  | 1.45 ± 0.27 |  | 4.44 ± 0.43 |  | 0.02 ± 0.0 |  | 96.28 ± 0.36 |  | 95.57 ± 0.45 |
| Quantitative evaluation metrics: All values are in %, presented as mean ± standard deviation. SSAT = Superficial Subcutaneous Adipose Tissue, DSAT = Deep Subcutaneous Adipose Tissue, IPAT = Intraperitoneal Adipose Tissue, RPAT = Retroperitoneal Adipose Tissue, PSAT = Paraspinal Adipose Tissue, Dice = Dice Similarity Coefficient, FP = False Positive Rate, FN = False Negative Rate. | | | | | | | | | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table S2. Volumetric comparison between ground truth and model 1 predictions.** | | | | | | |
| **Fat Depot** | **Hold-out test set (S-PRESTO, N=89)** | | | **External test set (SAMS, N=50)** | | |
| **Ground Truth (cm3)** | **Prediction (cm3)** | **Mean diff./ GT (%)** | **Ground Truth (cm3)** | **Prediction (cm3)** | **Mean diff./ GT (%)** |
| **SSAT** | 1627.3  (1197.4, 2360.3) | 1660.4  (1200.1, 2349.6) | 0.78 | 1303.3  (953.3, 1750.5) | 1354.9  (939.6, 1800.0) | 3.14 |
| **DSAT** | 1084.0  (777.2, 1533.6) | 1072.3  (781.3, 1520.6) | 1.40 | 879.0  (576.1, 1169.0) | 849.3  (547.5,1099.5) | 4.67 |
| **IPAT** | 441.9  (276.0, 825.2) | 433.91  (277.1, 829.3) | 0.21 | 623.0  (310.4, 1082.2) | 665.7  (311.6,1091.5) | 0.29 |
| **RPAT** | 333.7  (238.0, 451.8) | 336.5  (239.8, 239.8) | 0.04 | 431.3  (292.1, 636.2) | 441.9  (300.8,643.0) | 0.56 |
| **PSAT** | 109.6  (91.6, 125.9) | 109.5  (91.4, 126.12) | 0.66 | 84.4  (69.1, 104.3) | 71.6  (63.0, 88.6) | 14.66 |
| Volumes are presented in cm3, median (Q1, Q3), Mean diff. = Mean difference between ground truth (GT) and predicted volumes, SSAT=Superficial Subcutaneous Adipose Tissue, DSAT=Deep Subcutaneous Adipose Tissue, IPAT=Intraperitoneal Adipose Tissue, RPAT=Retroperitoneal Adipose Tissue, PSAT=Paraspinal Adipose Tissue | | | | | | |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table S3. Mean evaluation results over 5-fold cross-validation with and without data augmentation** | | | | | | | | | | | |
| **Fat depot** | **Augmen-tation** | **Hold-out test set (S-PRESTO, N=98)** | | | |  | **External test set (SAMS, N=50)** | | | | |
| **Dice**  **(%)** | **HD95**  **(mm)** | **Precision (%)** | **Sensitivity**  **(%)** |  | **Dice**  **(%)** | **HD95**  **(mm)** | **Precision (%)** | **Sensitivity**  **(%)** |
| **SSAT** | **Yes**  **No** | 98.21 ± 0.05  98.17 ± 0.02 | 1.78 ± 0.86  1.79 ± 0.90 | 98.15 ± 0.15  98.11 ± 0.14 | 98.19 ± 0.16  98.17 ± 0.18 |  | 95.66 ± 3.61  95.73 ± 2.61 | 2.07 ± 1.62  3.02 ± 4.45 | 95.66 ± 4.22  94.05 ± 3.6 | 95.78 ± 4.14  97.54 ± 2.62 |
| **DSAT** | **Yes**  **No** | 97.29 ± 0.07  97.26 ± 0.02 | 1.80 ± 0.90  1.83 ± 0.94 | 97.38 ± 0.37  97.26 ± 0.21 | 97.05 ± 0.32  96.88 ± 0.34 |  | 94.54 ± 3.64  93.37 ± 5.49 | 3.48 ± 10.41  3.49 ± 5.71 | 94.46 ± 4.42  94.05 ± 7.64 | 94.72 ± 3.80  93.06 ± 4.19 |
| **IPAT** | **Yes**  **No** | 96.76 ± 0.15  96.75 ± 0.14 | 1.54 ± 0.24  1.53 ± 0.32 | 96.91 ± 0.23  96.79 ± 0.29 | 97.22 ± 0.32  97.22 ± 0.08 |  | 93.45 ± 4.16  92.62 ± 4.57 | 2.29 ± 2.62  2.82 ± 3.30 | 94.48 ± 4.89  90.49 ± 6.36 | 92.57 ± 4.55  95.03 ± 3.57 |
| **RPAT** | **Yes**  **No** | 96.50 ± 0.09  96.43 ± 0.11 | 1.54 ± 0.25  1.57 ± 0.38 | 96.53 ± 0.26  96.37 ± 0.19 | 96.73 ± 0.20  96.74 ± 0.12 |  | 92.71 ± 4.26  92.07 ± 4.31 | 3.71 ± 4.38  4.63 ± 5.59 | 94.24 ± 2.76  92.66 ± 4.35 | 91.36 ± 5.97  91.54 ± 4.72 |
| **PSAT** | **Yes**  **No** | 96.08 ± 0.19  96.10 ± 0.26 | 1.22 ± 0.59  1.13 ± 0.65 | 96.41 ± 0.38  96.58 ± 0.19 | 95.78 ± 0.39  95.70 ± 0.47 |  | 88.49 ± 6.02  85.21 ± 6.73 | 3.52 ± 2.59  4.05 ± 4.3 | 91.88 ± 3.71  88.55 ± 4.97 | 85.86 ± 9.05  82.51 ± 9.34 |
| Quantitative evaluation metrics: All values are in %, presented as mean ± standard deviation. SSAT = Superficial Subcutaneous Adipose Tissue, DSAT = Deep Subcutaneous Adipose Tissue, IPAT = Intraperitoneal Adipose Tissue, RPAT = Retroperitoneal Adipose Tissue, PSAT = Paraspinal Adipose Tissue, Dice=Dice Similarity Coefficient. | | | | | | | | | | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table S4. Metrics definition** | | | | |
| **Metric** |  | **Definition** |  | **With:** |
| Precision |  |  |  | * TP=True positive * FP=False positive * FN=False negative |
| Sensitivity |  |  |  |
| Dice similarity coefficients (Dice) |  |  |  | * G=ground truth segmentation * P=predicted segmentation |
| Hausdorff distance 95th percentile (HD95) |  |  |  | * = |